Conhecimento e Pré-Processamento

Gabriel Fernandes Silva

Renomear colunas

O primeiro passo foi renomear as colunas para facilitar o entendimento dos dados

```
dicionario = {
    "Natureza":"Fim Lucrativo",
    "%": "Cursos Sem Ato 5 anos",
    "CI N": "Ultimo CI",
    "CI V": "Penultimo CI",
    "IGC N": "Ultimo IGC",
    "IGC M": "Penultimo IGC",
    "IGC V": "Antepenultimo IGC",
    "variacao mat": "Variacao Matricula 16/17",
    "mat T 2017": "Matriculas 17",
    "saldo 2017": "Saldo 2017",
    "Variacao 17-16": "Variacao do Saldo 16/17"
}

dados = dados.rename(columns=dicionario)
```

Valores Ausentes

Algumas colunas apresentavam dados ausentes como S/C ou S/D.

Esses valores foram substituidos por NaN com intuito de analisar a representatividade deles nos dados

Descrição dos Dados

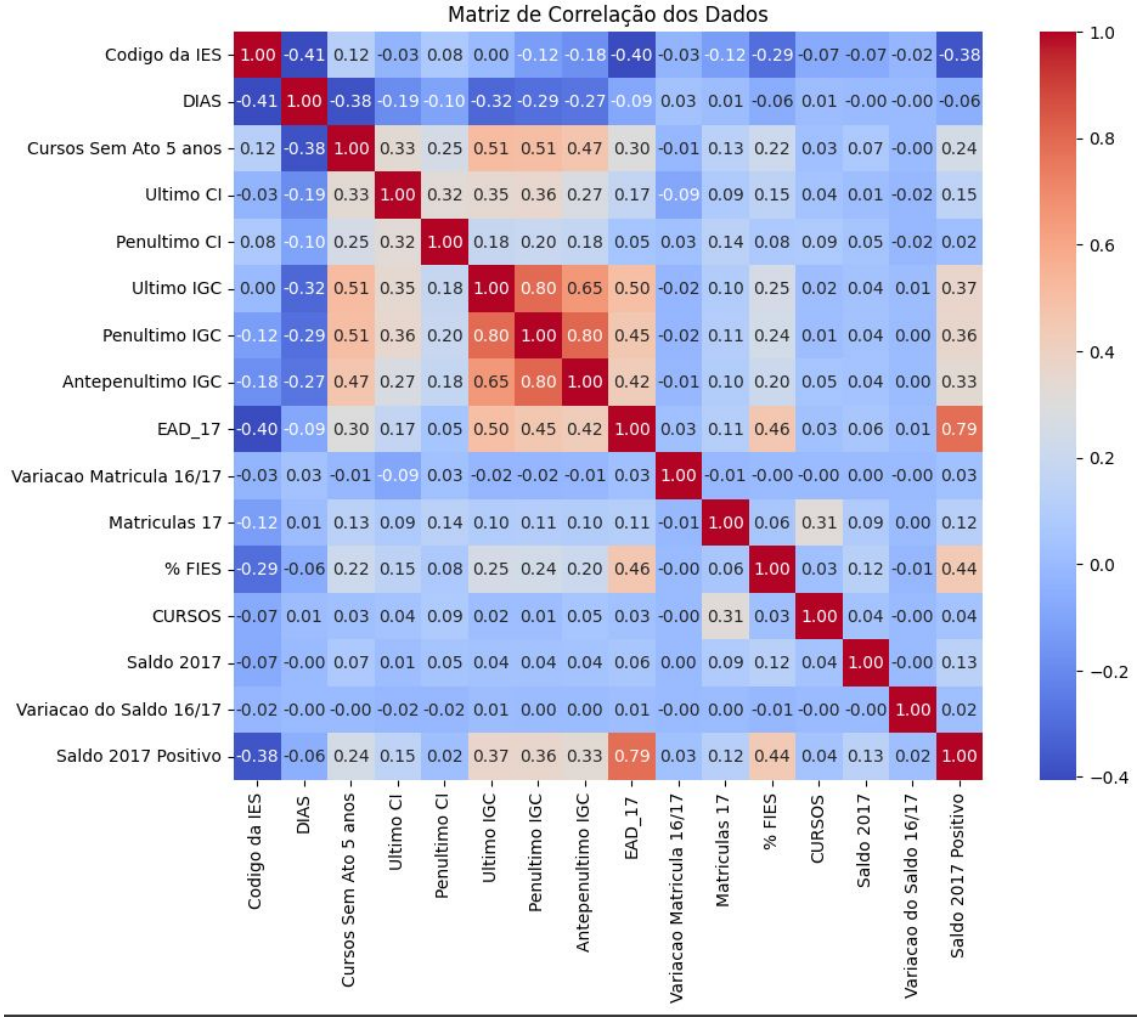
Penultimo CI - faltante

```
 0   Codigo da IES              2177 non-null   int64
 1   Fim Lucrativo              2177 non-null   int64
 2   Situacao                   2177 non-null   object
 3   DIAS                       2177 non-null   int64
 4   Cursos Sem Ato 5 anos      2177 non-null   float64
 5   Ultimo CI                  1733 non-null   object
 6   Penultimo CI               992 non-null    object
 7   Ultimo IGC                 1540 non-null   object
 8   Penultimo IGC              1541 non-null   object
 9   Antepenultimo IGC          1386 non-null   object
 10  EAD_17                     2177 non-null   int64
 11  Variacao Matricula 16/17   2177 non-null   float64
 12  Matriculas 17              2177 non-null   int64
 13  % FIES                     2177 non-null   float64
 14  CURSOS                     2177 non-null   int64
 15  Saldo 2017                 2177 non-null   float64
 16  Variacao do Saldo 16/17    2177 non-null   float64
dtypes: float64(5), int64(6), object(6)
```

Correlação

IGC - Apresentou uma correlação forte entre o último, penúltimo e antepenúltimo.

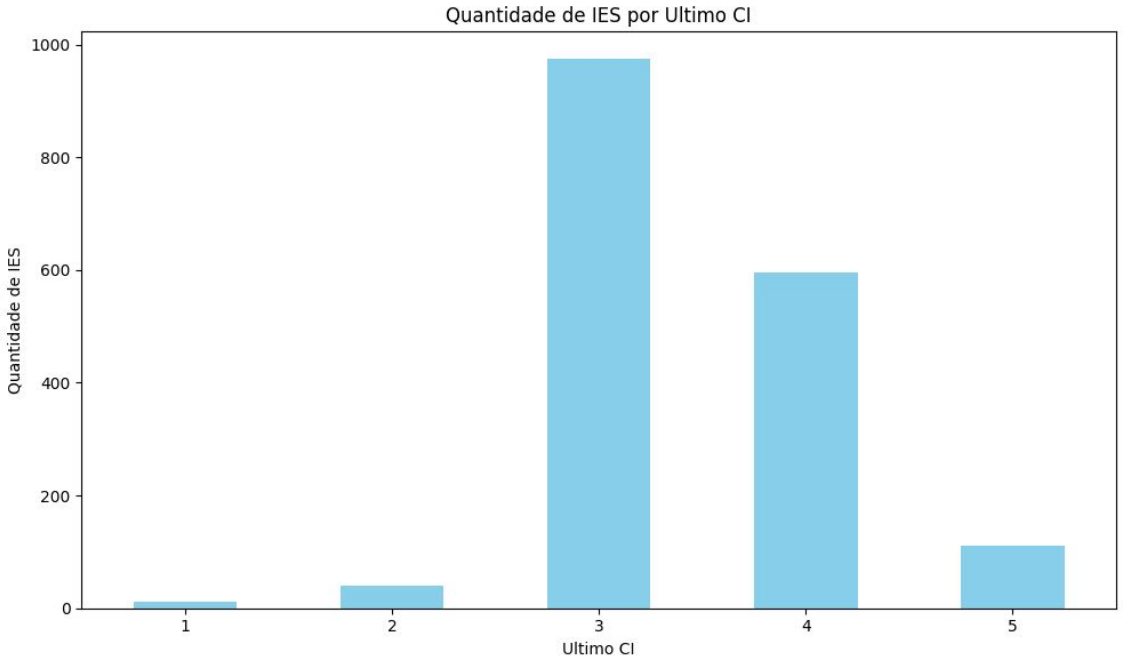
EAD - Saldo 2017 Positivo, apresentou uma correlação, justificável pelo modelo EAD ser mais escalável

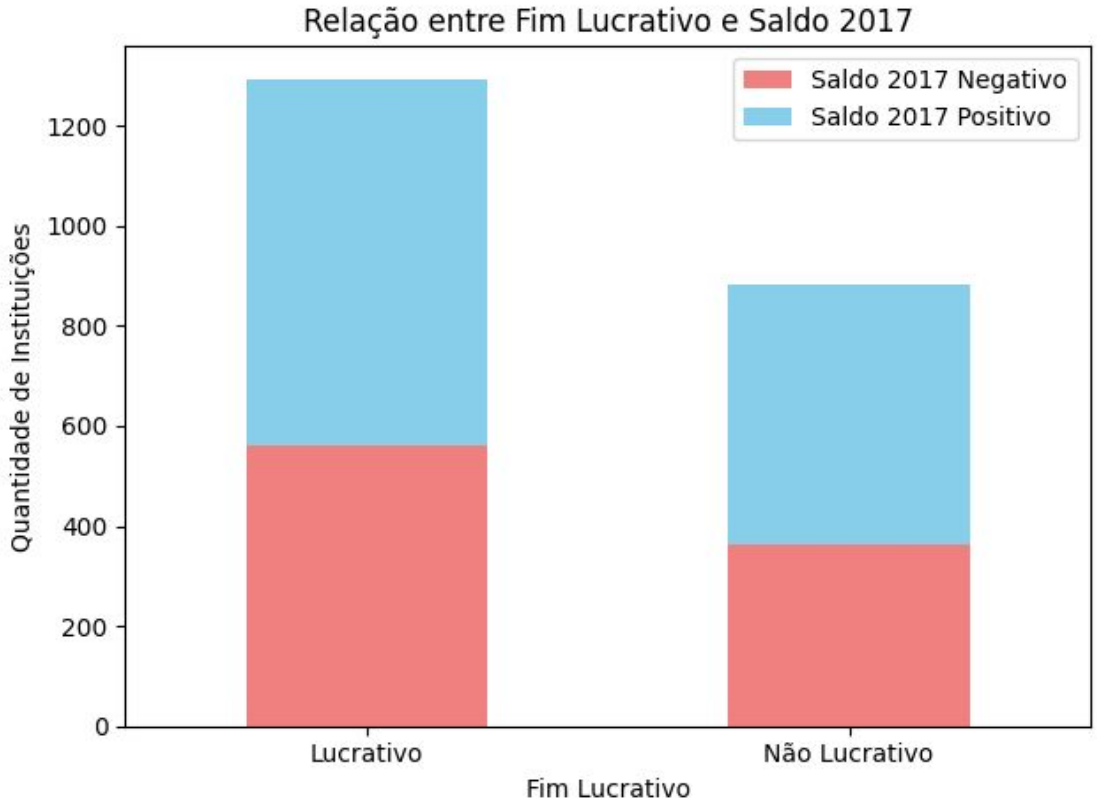


IES em relação ao CI

Maior parte das IES estão com notas 3 e 4.

Distribuição não é muito normalizada





Relação de resultado com a finalidade da instituição

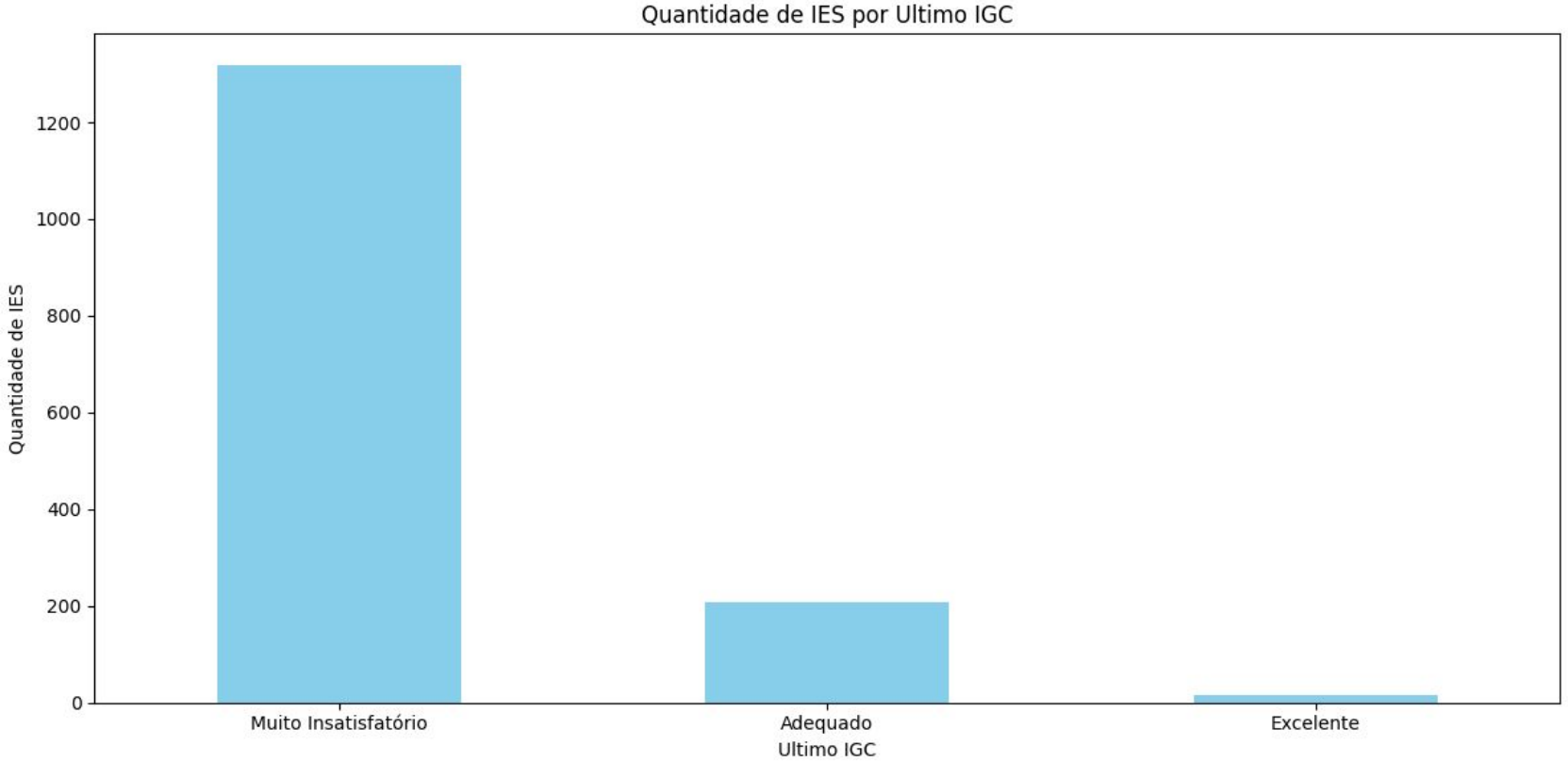
IGC

Indice Geral de Cursos

0 - 3 -> Muito Insatisfatório

3 - 4 -> Adequado

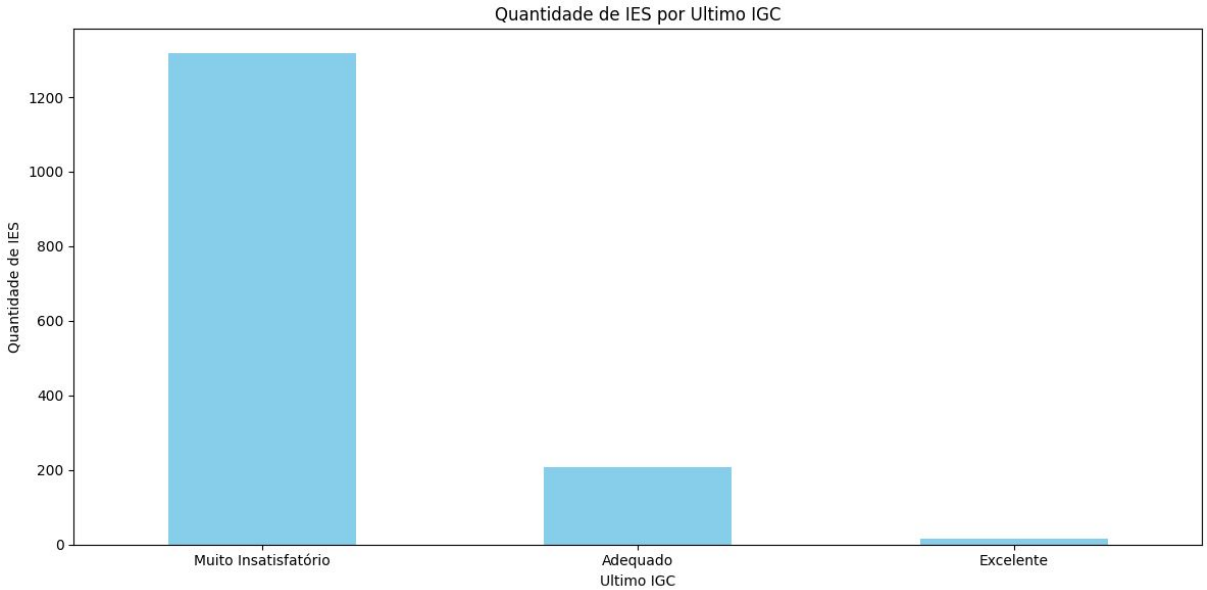
4 - 5 -> Excelente



Análise categórica do IGC

IGC

Diferente do CI, o IGC apresentou um comportamento bem mais tendencioso para o cenário Insatisfatório (<3)



Limpeza

Remover colunas redundantes e dados faltantes não estimáveis

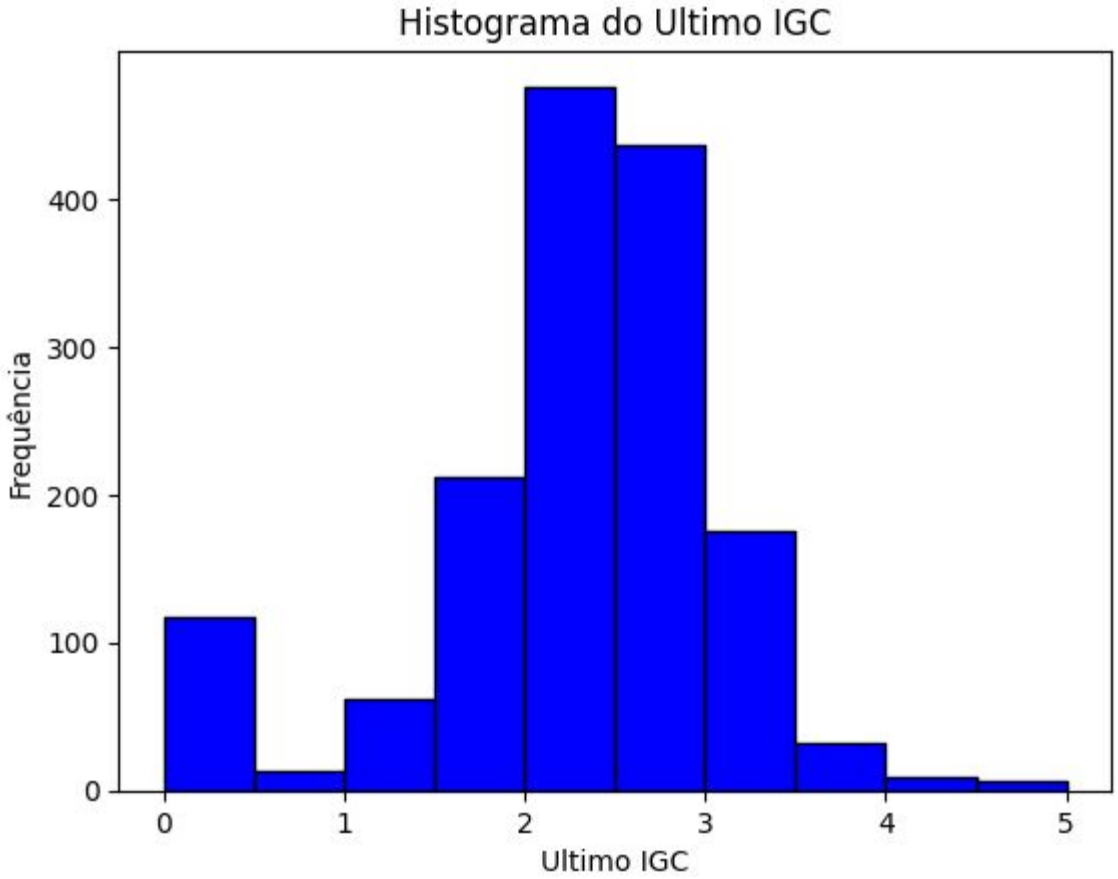
IGC

Apresentou uma correlação forte entre último, penúltimo e antepenúltimo.

E alguns dados estavam inconsistentes, como valores negativos e valores acima de 5

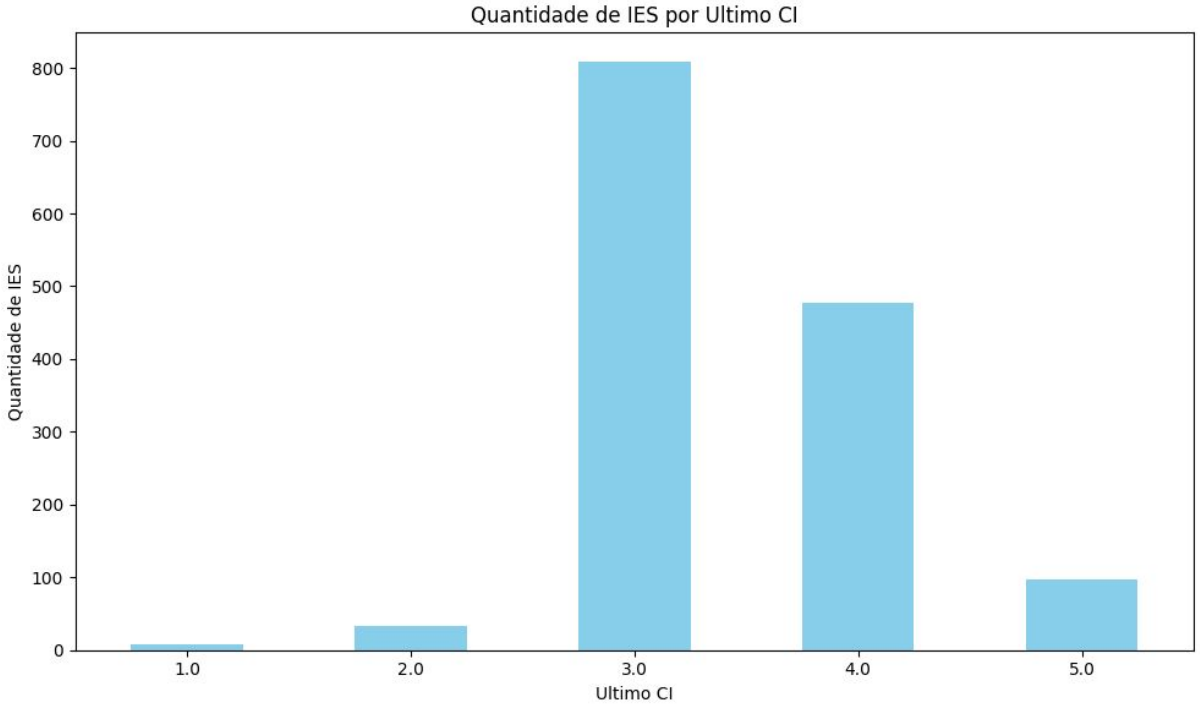
Os negativos foram excluídos, pois representavam uma amostra muito reduzida (~8)

Os valores maiores que 5 também representavam uma amostra reduzida



CI

O penúltimo CI possui uma quantidade pequena de dados, então a coluna foi excluída e o último CI foi usado como parâmetro de CI

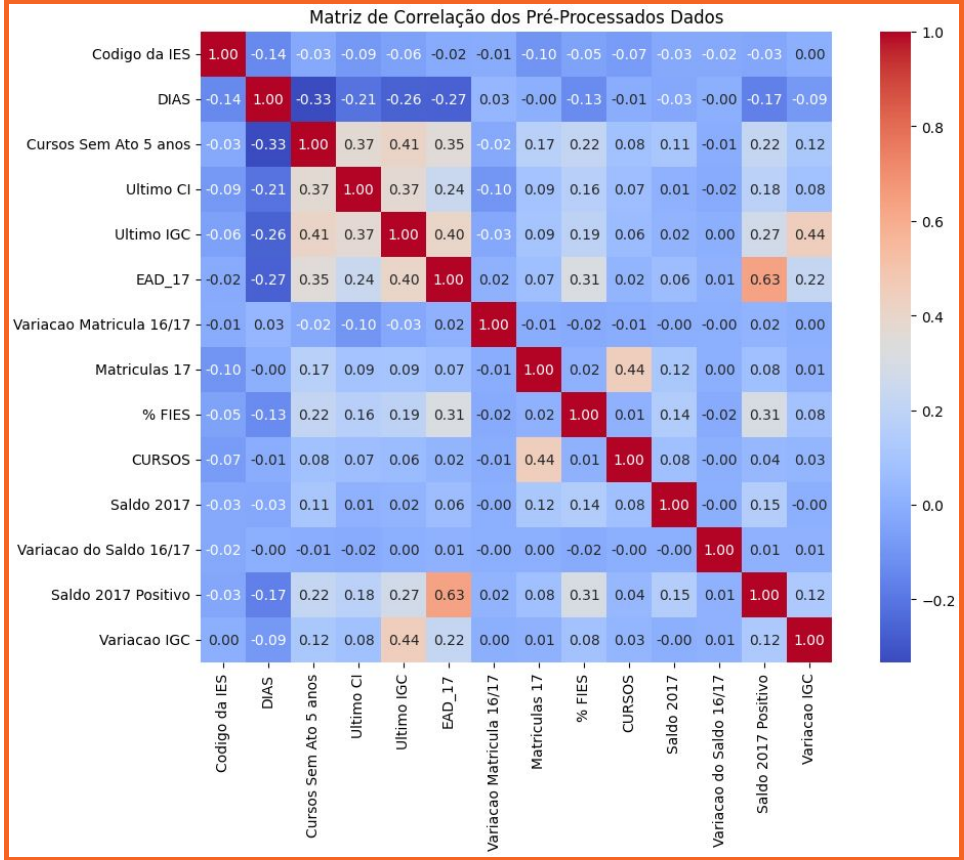


Descrição dos Dados Pré-Processados

```
 0   Codigo da IES              1424 non-null   int64
 1   Fim Lucrativo              1424 non-null   category
 2   Situacao                   1424 non-null   category
 3   DIAS                       1424 non-null   int64
 4   Cursos Sem Ato 5 anos      1424 non-null   float64
 5   Ultimo CI                  1424 non-null   float64
 6   Ultimo IGC                 1424 non-null   float64
 7   EAD_17                     1424 non-null   int64
 8   Variacao Matricula 16/17   1424 non-null   float64
 9   Matriculas 17              1424 non-null   int64
 10  % FIES                     1424 non-null   float64
 11  CURSOS                     1424 non-null   int64
 12  Saldo 2017                 1424 non-null   float64
 13  Variacao do Saldo 16/17    1424 non-null   float64
 14  Saldo 2017 Positivo        1424 non-null   bool
 15  Ultimo IGC Categoria       1424 non-null   category
 16  Variacao IGC               1424 non-null   float64
dtypes: bool(1), category(3), float64(8), int64(5)
```

Correlação

Pré-Processados



Obrigado!